

DMDC TR 99-05

Evaluating Large-Scale Training Simulations

Volume I: Reference Manual

Henry Simpson

Defense Manpower Data Center • Seaside, California

REPORT DOCUMENTATION PAGE			<i>Form Approved</i> OMB No. 074-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE December 1999	3. REPORT TYPE AND DATES COVERED Final		
4. TITLE AND SUBTITLE EVALUATING LARGE-SCALE TRAINING SIMULATIONS, VOLUME I: REFERENCE MANUAL		5. FUNDING NUMBERS		
6. AUTHOR(S) Henry Simpson				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Defense Manpower Data Center DoD Center, Monterey Bay 400 Gigling Road Seaside, CA 93955-6771		8. PERFORMING ORGANIZATION REPORT NUMBER DMDC Technical Report 99-05		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Deputy Under Secretary of Defense (Readiness) 4000 Defense, The Pentagon Washington, DC 20301-4000		10. SPONSORING / MONITORING AGENCY REPORT NUMBER		
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 Words) <p>Objectives of the manual are to (1) provide guidance to help analysts design meaningful training effectiveness evaluations, (2) describe procedures for alternative methods of conducting training effectiveness evaluations, and (3) provide examples of training effectiveness evaluations that may be used as models to emulate. Chapter 1 (Introduction) describes the problem and issues, objectives, and method. Chapter 2 (Building an Evaluation Framework) explains why people conduct evaluations. Chapter 3 (Evaluation Methods) describes evaluation methods and provides examples of their application. Chapter 4 (Case Studies) describes well-documented evaluations: SIMNET/CCTT (Simulation Networking/Close Combat Tactical Trainer) and MDT2 (Multi-service Distributed Training Testbed). Chapter 5 (Evaluation Problem Areas) contrasts laboratory and field evaluations, discusses lessons learned from past evaluations, and critiques field evaluation practice. Chapter 6 (Procedural Guidance) identifies and summarizes published evaluation guidance. Chapter 7 (Evaluation Criteria) discusses how evaluation criteria differ depending upon evaluation method, for small- and large-scale evaluations, and depending upon evaluation perspective (training versus system developer versus modeling and simulation). Chapter 8 (Evaluation Framework) presents the evaluation framework in terms of evaluation principles and a description of the timing of evaluation events, their purpose, and relevant dependent variables-linked to relevant examples and procedural guidance.</p>				
14. SUBJECT TERMS large-scale training simulation, training evaluation, evaluation methods, training effectiveness, military training			15. NUMBER OF PAGES 190	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

To Jesse Orlansky

Preface

The Office of the Inspector General, Department of Defense, recommended that the Under Secretary of Defense for Personnel and Readiness establish policy and procedures to evaluate the training and cost-effectiveness of large-scale training simulations. One of the Under Secretary's responses to this request was to direct the Defense Manpower Data Center to develop guidelines to help evaluators conduct evaluations and to develop a historical training effectiveness data base. This volume describes the work performed by Defense Manpower Data Center in response to the Under Secretary's request and the resulting products and actions: Guidance to help evaluators design meaningful training effectiveness evaluations, descriptions of procedures for alternative methods, methodological examples, historical training effectiveness data base, and user access to the data base. These products and actions are intended to help the military Services determine when and how to evaluate the training and cost-effectiveness of large-scale training simulations.

This manual consists of two volumes: I (Reference Manual) and II (User's Manual). Volume II is designed to give readers a quick start introduction to training evaluation and a roadmap to the more in-depth content of Volume I. Readers are encouraged to start with Volume II.

The views expressed in this manual are those of the author, are not official, and do not necessarily reflect an official policy position of the Services, the Department of Defense, or the U.S. Government.

Acknowledgments

Dan Gardner sponsored the development of this manual. I thank him for his encouragement and patience.

More than two dozen individuals contributed ideas to this manual. The manual also uses many case studies based on 250 training evaluations conducted by literally hundreds of researchers. I have attempted to knit the ideas and evaluation practices into an overall evaluative framework for large-scale training simulations. The names of the many influences on this manual are too numerous to list here but will be evident from the citations.

Fred Hartman facilitated the review of early drafts of this manual. I thank him for providing many suggestions for improving the manual.

Thanks to Richard Kass for allowing me to reproduce his *Test Officer's Guide for Designing Valid Tests and Experiments* in Chapter 6.

Randy Marks of DMDC designed and composed the manual with the assistance of Aaron Toy, who created the illustration for the cover .

Dan Gardner, Fred Hartman, Jack Leather, and John Morrison reviewed this manual. Herbert Bell, David Bessemer, Edward George, Jack Hiller, Richard Kass, Douglas Macpherson, Angelo Mirabella, Frank Moses, Randy Oser, Robin Rose, and Uldi Shvern reviewed portions of the manual describing the evaluation framework. Organizational affiliations are shown below.

The individuals listed on the facing page provided information, opinions, documents, and in other ways supported the development of the manual. The presence of a name here or above does not indicate endorsement of the manual. I assume full responsibility for its contents and any errors, omissions, or oversights.

Henry Simpson
Seaside, California

Dee Andrews—Air Force Armstrong Laboratory
Herbert Bell—Air Force Armstrong Laboratory
David W. Bessemer—Army Research Institute
John Boldovici—Army Research Institute
Janis A. Cannon-Bowers—Naval Air Warfare Center Training Systems Division
Kenneth D. Cross—Bayview Research
Philip Djang—TRADOC Analysis Center
John Ellis—Navy Personnel Research and Development Center
Dorothy Finley—Army Research Institute
Dan Gardner—Office of the Deputy Under Secretary of Defense for Readiness
Edward L. George—TRADOC Analysis Center
Joseph Hagman—Army Research Institute
Fred Hartman—Institute for Defense Analyses
John Hayes—Army Research Institute
Jack H. Hiller—Army Research Institute
Donald Johnson—Office of the Deputy Under Secretary of Defense for Readiness
Richard Kass—Army Test and Experimentation Command
Peter Kincaid—University of Central Florida Institute for Simulation and Training
Richard Laferriere—TRADOC Analysis Center
Jack Leather—Defense Manpower Data Center
Douglas Macpherson—Army Research Institute
Angelo Mirabella—Army Research Institute
Franklin L. Moses—Army Research Institute
Randy Oser—Naval Air Warfare Center Training Systems Division
Jesse Orlansky¹—Institute for Defense Analyses
Ruth Phelps—Army Research Institute
William Rankin—Naval Air Warfare Center Training Systems Division
J. Wesley Regian—Air Force Armstrong Laboratory
Robin Rose—TRADOC Analysis Center
Eduardo Salas—Naval Air Warfare Center Training Systems Division
Robert Seidel—Army Research Institute
Uldi Shvern—Army Operational Test and Experimentation Command
Henry L. Taylor—Institute of Aviation, University of Illinois
Diana Tierney—Deputy Chief of Staff for Training, TRADOC
Robert Worley—Institute for Defense Analyses

¹ Deceased.

EXECUTIVE SUMMARY

Problem and Issues

The Department of Defense Office of the Inspector General (DoDIG) conducted an audit concerning the impact on readiness of training simulators and devices. The audit focused particular attention on shortcomings in evaluation of large-scale training simulations (LSTS). The DoDIG recommended that the Under Secretary of Defense for Personnel and Readiness establish policy and procedures to evaluate the training and cost-effectiveness of LSTS. One of the Under Secretary's responses to this request was to direct the Defense Manpower Data Center (DMDC) to develop guidelines to help evaluators conduct evaluations and a historical training effectiveness data base. This manual describes the work performed by DMDC and the resulting products and actions: Guidance to help evaluators design meaningful training effectiveness evaluations, descriptions of procedures for alternative methods, methodological examples, historical training effectiveness data base, and user access to the data base. These products and actions are intended to help the military Services determine when and how to evaluate the training and cost-effectiveness of LSTS.

Objectives

Objectives of this manual are to:

- Provide guidance to help analysts design meaningful training effectiveness evaluations.
- Describe procedures for alternative methods of conducting training effectiveness evaluations.
- Provide examples of training effectiveness evaluations that may be used as models to emulate.

Method

The method consisted of literature review, development of a historical training and cost-effectiveness data base, analyses, development of guidelines, identification of case studies, and review of findings by subject-matter experts.

Evaluation Guidance

This manual contains evaluation guidance in eight chapters and two appendices. Chapter 1 (Introduction) describes the problem and issues, objectives, method, and shows where the manual addresses its objectives. Chapter 2 (Building an Evaluation Framework) explains why people conduct training effectiveness evaluations and starts to build an evaluation framework for LSTS by attempting to answer fundamental questions about the evaluation process (Whose training is evaluated? What is evaluated? Where to evaluate? How to evaluate? What are evaluation criteria? When to evaluate?). Chapter 3 (Evaluation Methods) describes the methods commonly used in military training effectiveness evaluations and provides examples of their application. Chapter 4 (Case Studies) reviews two well-documented evaluations of LSTS: SIMNET/CCTT (Simulation Networking/Close Combat Tactical Trainer) and MDT2 (Multi-Service Distributed Training Testbed). Chapter 5 (Evaluation Problem Areas) contrasts laboratory and field evaluations, discusses lessons learned from past evaluations, and critiques field evaluation practice. Chapter 6 (Procedural Guidance) identifies and summarizes published evaluation guidance. Chapter 7 (Evaluation Criteria) discusses how evaluation criteria differ depending upon evaluation method, for small- and large-scale evaluations, and perspective. Chapter 8 (Evaluation Framework) presents the evaluation framework. Appendix A (Reference Lists for Chapter 3) contains reference lists. Appendix B (Acronyms) defines acronyms. Author and Subject indexes are provided.

C O N T E N T S

CHAPTER 1. INTRODUCTION	1
Overview	1
Problem and Issues	2
Objectives	4
Method	4
Road Map: Where This Guide Addresses Each of Its Objectives	7
CHAPTER 2. BUILDING AN EVALUATION FRAMEWORK.....	9
Why Evaluate?	10
Whose Training Is Evaluated?	12
What Is Evaluated?	14
Where To Evaluate?.....	16
How To Evaluate?.....	16
What Are Evaluation Criteria?	21
When To Evaluate?.....	22
CHAPTER 3. EVALUATION METHODS.....	25
Experiments	25
Judgment-Based Evaluations.....	36
Analytical Evaluations	40
Surveys	47
CHAPTER 4. CASE STUDIES	49
Finding Cases	49
Two Good Cases.....	50
SIMNET/CCTT	50
MDT2.....	58
Reference List 4-1. SIMNET/CCTT Publications.....	67
Reference List 4-2. MDT2 Publications	70
CHAPTER 5. EVALUATION PROBLEM AREAS	73
Field Versus Laboratory Evaluations	73
Lessons Learned	77
Critiques of Field Evaluation Practice	80
CHAPTER 6. PROCEDURAL GUIDANCE	85
Evaluation Methods.....	85
System and Program Evaluation Frameworks.....	93
Collective and Team Training	95
Reference List 6-1: Procedural Guidance.....	97
CHAPTER 7. EVALUATION CRITERIA	101
Defining Evaluation Criteria.....	101
Evaluation Criteria for an LSTS Evaluation	106
CHAPTER 8. EVALUATION FRAMEWORK	111
Evaluation Objectives.....	111
Evaluation Principles	113
Evaluation Framework	116
Reference List 8-1. Evaluation Framework.....	131

REFERENCES	137
APPENDIX A. REFERENCE LISTS FOR CHAPTER 3	155
Overview.....	155
Reference List A-1. True Experiment.....	155
Reference List A-2. Pre-Experiment.....	161
Reference List A-3. Quasi-Experiment.....	163
Reference List A-4. Test.....	164
Reference List A-5. Transfer Experiment.....	166
Reference List A-6. Ex Post Facto	167
Reference List A-7. Judgment (Users)	169
Reference List A-8. Judgment (SMEs).....	170
Reference List A-9. Judgment (Analysts)	171
Reference List A-10. Analysis (Evaluate)	172
Reference List A-11. Analysis (Compare).....	174
Reference List A-12. Analysis (Optimize)	175
Reference List A-13. Survey	176
APPENDIX B. ACRONYMS.....	179
AUTHOR INDEX.....	183
SUBJECT INDEX.....	187
FIGURES	
4-1 TARGET's data for planning phase (from Orlansky et al., 1997)	63
4-2 TOM data for planning phase (from Orlansky et al., 1997).....	64
5-1 Notional representation of a simple inverse relationship between internal and external validity for laboratory and field studies (adapted from Shute and Regian, 1993).	76
6-1 Test Officer's Guide for Designing Valid Tests and Experiments (from Kass, 1997, reproduced courtesy of the author).....	88
TABLES	
2-1 Some Key Evaluation Questions.....	10
2-2 Training Content, Echelon, and Training Versus Education Taxonomy.....	13
2-3 Training Type and Subtype Taxonomy	15
2-4 Frequency of Usage of Common Evaluation Methods and Submethods	19
2-5 Levels of Data Commonly Associated with Evaluation Methods	21
3-1 Three True Experimental Designs (from Campbell & Stanley, 1966)	26
3-2 Three Pre-Experimental Designs (from Campbell & Stanley, 1966)	28
3-3 Ten Quasi-Experimental Designs (from Campbell & Stanley, 1966).....	30
3-4 Three Classes of Transfer Designs (adapted from Pfeiffer & Browning, 1984).....	34
3-5 Frequencies and Percentages of Usage of Analyst-, SME-, and User Judgment-Based Evaluations in TCEF Sample	38
3-6 Frequencies and Percentages of Usage of Three Classes of Analysis (Evaluate, Compare, or Optimize) for Existing Versus Hypothetical Systems.	41

TABLES (continued)

4-1	SIMNET/CCTT Evaluations by Authors and Evaluation Method and Submethod	54
4-2	Simulators, Service, Types of Participant, and Locations Linked in MDT2 (adapted from Orlansky et al., 1997).....	60
4-3	Dimensions and Subdimensions of TOM (Teamwork Observation Measure) (adapted from Orlansky et al., 1997).....	62
4-4	Bombing Performance for February 1995 Exercise (from Orlansky et al., 1997).....	64
4-5	CAS Kills, Misses, and Average Engagement Time for February 1995 Exercise (adapted from Orlansky et al., 1997).....	64
4-6	Results of MDT2 Survey (from Orlansky et al., 1997).	65
5-1	Potential Limitations of Reported Findings on Training Device Effectiveness (Based on N=35 evaluations) (adapted from Kraemer & Rowatt, 1993)	83
7-1	Dependent Variables Used in SIMNET/CCTT Evaluations by Authors and Evaluation Methods	102
7-2	Consolidated List of Recommended Dependent Measures by Author.....	108
7-3	Consolidated List of Recommended Dependent Measures with Descriptions	108
8-1	Common Objectives for Conducting Training Effectiveness Evaluations	112
8-2	Descriptive Summary of Representative Evaluations for Evaluation Objective 2A: Predict Training Effectiveness	118
8-3	Descriptive Summary of Representative Evaluations for Evaluation Objective 2B: Measure Training Effectiveness.....	119
8-4	Evaluation Criteria by Study for Objective 2B: Measure Training Effectiveness	119
8-5	Descriptive Summary of Representative Evaluations for Evaluation Objective 2C: Determine Training Effectiveness	121
8-6	Descriptive Summary of Representative Evaluations for Evaluation Objective 3B: Measure Transfer of Training	123
8-7	Evaluation Criteria by Study for Objective 3B: Measure Transfer of Training	123
8-8	Descriptive Summary of Representative Evaluations for Evaluation Objective 4B (Measure User Acceptance) and 4C (Determine User Acceptance).....	127
8-9	Descriptive Summary of Representative Evaluations for Evaluation Objective 5(AB): Support Training Design	128
8-10	Descriptive Summary of Representative Evaluations for Evaluation Objective 5C: Determine Training Status.....	129

